Online tests of Kalman filter consistency

Online tests of Kalman filter consistency

Robert Piché
Department of Automation Science and Engineering
Tampere University of Technology
Tampere, Finland

December 9, 2014

Abstract

The normalised innovation squared (NIS) test, which is used to assess whether a Kalman filter's noise assumptions are consistent with realised measurements, can be applied online with real data, and does not require future data, repeated experiments, or knowledge of the true state. In this work it is shown that the NIS test is equivalent to three other model criticism procedures: it can be derived as a Bayesian p-test for the prior predictive distribution, as a nested-model parameter significance test, and from a recently-proposed filter residual test. A new NIS-like test corresponding to a posterior predictive Bayesian p-test is presented.

1 Introduction

If the measurement noise covariance parameter in a Kalman filter is too small relative to the actual noise, the filter gives too much weight to measurements relative to the process model, and estimated state trajectories are overly erratic. On the other hand, if the parameter is too large, the filter gives too little weight to measurements, and its response is sluggish.

Bar-Shalom et al. [1] have proposed a statistical hypothesis test for detection of model mismatch (also called filter inconsistency). An acceptance region for the test is defined based on the fact that, under the hypothesis that the filter model is correct, the "normalised innovation squared" (NIS)

statistic has a chi-square distribution. The test can be applied online because the NIS statistic can be computed from current and past measurements and does not use future measurements or knowledge of the true state. Because measurements at a single time step may not give enough information for reliable assessment of consistency, Bar-Shalom et al. recommend that the test be applied to the average of NIS values from several consecutive time steps.

The filter residual test recently proposed by Gibbs [2] is also feasible for online implementation. This test is formulated in terms of individual components of the residual and is intended for detection of measurement outliers, but it is straightforward to adapt this approach to derive a sum of squared residuals statistic in a test of filter consistency. As shown in section 2.2, the sum of squared residuals statistic is in fact equivalent to NIS.

The aim of this study is to obtain a better understanding of existing online filter consistency testing and to develop new methods, by adapting methods of model criticism from Bayesian statistical theory to the Kalman filtering setting. Bayesian model criticism is an active research area and many approaches have been proposed; see chapter 8 of [3] for a survey. Here, three approaches will be considered; two of these turn out be be equivalent to the NIS test, while one leads to a new NIS-like filter consistency test.

One basic approach to model criticism is to assess whether the realised measurements are "surprising" according to the assumed statistical model. Box [4] suggests that this surprise be quantified using p-value diagnostics based on the prior predictive distribution. In section 2.2 it is shown that the NIS test can be interpreted as a Bayesian prior predictive p-value diagnostic.

Gelman et al. [5] propose p-value diagnostics based on the *posterior* predictive distribution. Although this approach makes redundant use of information (the surprisingness of data is evaluated using a distribution that was constructed using the data), it is argued to be more effective than the prior predictive p-test especially in cases when the prior is very diffuse. A posterior predictive variant of the NIS test is presented in section 2.2.

A more fundamental objection to Bayesian p-values is its use of hypothesis testing, a procedure from classical statistics. An alternative approach that is more consistent with Bayesian statistical theory is advocated by Kruschke [6]. He constructs an augmented statistical model in which constants in the base model are treated as additional unknown parameters to be estimated. Model misfit is diagnosed if the marginal posterior distribution of the additional parameters has most of its probability far from the nominal values that correspond to the base model. In section 2.3 it is shown how the

NIS test can be derived using an augmented model approach.

Section 3 presents simulation studies of the performance of the NIS test and its posterior predictive variant in a basic benchmark example. A theoretical study of the consistency tests applied to a Kalman filter for multiple measurements of a scalar stationary state is presented in section 4.

2 Kalman Filter Consistency Tests

2.1 Base model

Consider the Kalman filter's measurement update stage, where the predicted state[1] x (i.e. the state after propagation through the process model) is combined with the measurement y via Bayes' theorem to produce the updated state $x \,|\, y$. The predicted state has a multivariate normal distribution with mean μ_0 and covariance P_0; this is denoted $x \sim \mathrm{MVN}(\mu_0, P_0)$. The sampling model for the n-variate measurement is $y \,|\, x \sim \mathrm{MVN}(Hx, R)$. Then by Bayes' theorem, and introducing the notation

$$S_0 = HP_0H' + R, \tag{1a}$$
$$K = P_0H'S_0^{-1}, \tag{1b}$$
$$P_1 = (I - KH)P_0, \tag{1c}$$
$$\mu_1 = \mu_0 + K(y - H\mu_0), \tag{1d}$$

the posterior distribution is

$$x \,|\, y \sim \mathrm{MVN}(\mu_1, P_1). \tag{2}$$

The distribution of a possible measurement replication $\tilde{y}$ that is conditionally independent of y given x is called a *predictive distribution*. In the Kalman filter update, the prior predictive distribution is

$$\tilde{y}_0 \sim \mathrm{MVN}(H\mu_0, S_0) \tag{3}$$

and the posterior predictive distribution is

$$\tilde{y}_1 \sim \mathrm{MVN}(H\mu_1, S_1), \tag{4}$$

[1] In filter theory, the predicted and updated states are conventionally denoted $x_{t|t-1}$ and $x_{t|t}$. For the sake of readability, time-index subscripts are omitted in this section.

where

$$S_1 = HP_1H' + R. \tag{5}$$

Formulas (2–4) are standard results of linear-Gaussian estimation theory; derivations are outlined in Appendix A.

2.2 Bayesian p-tests

Consider first a test statistic based on the prior predictive distribution. It follows from (3) that the random variable

$$(\tilde{y}_0 - H\mu_0)'S_0^{-1}(\tilde{y}_0 - H\mu_0) \tag{6}$$

has a χ^2_n distribution. A Bayesian p-test to assess whether the realised measurement y is "surprisingly" large or small with respect to the prior predictive distribution is to determine whether the corresponding test statistic

$$\epsilon_0 = (y - H\mu_0)'S_0^{-1}(y - H\mu_0) \tag{7}$$

lies in the tails of the χ^2_n distribution. This procedure is equivalent to the NIS test of [1, p. 236],

In Appendix A, the following equivalent formula for the NIS test statistic is derived:

$$\epsilon_0 = (\mu_1 - \mu_0)'P_0^{-1}(\mu_1 - \mu_0) + (y - H\mu_1)'R^{-1}(y - H\mu_1). \tag{8}$$

Instead of the prior predictive measurement $\tilde{y}_0$, Gibbs [2] proposes a test statistic based on the prior predictive residual $\tilde{y}_0 - H\tilde{\mu}_1$, where

$$\tilde{\mu}_1 = \mu_0 + K(\tilde{y}_0 - H\mu_0) \tag{9}$$

is the prior predictive state estimate. The prior predictive residual's distribution is

$$\tilde{y}_0 - H\tilde{\mu}_1 \sim \mathrm{MVN}(0, RS_0^{-1}R), \tag{10}$$

(see Appendix A), and so

$$(\tilde{y}_0 - H\tilde{\mu}_1)'R^{-1}S_0R^{-1}(\tilde{y}_0 - H\tilde{\mu}_1) \sim \chi^2_n. \tag{11}$$

The corresponding test statistic for a Bayesian p-test of the realised residual $y - H\mu_1$ with respect to its prior predictive distribution is then

$$\epsilon_r = (y - H\mu_1)'R^{-1}S_0R^{-1}(y - H\mu_1). \tag{12}$$

Substituting the identity $y - H\mu_1 = RS_0^{-1}(y - H\mu_0)$ into (12), it follows that $\epsilon_r = \epsilon_0$. That is, the normalised sum of squared residuals statistic is equal to the NIS statistic.

Because the number of measurements available at a single time step can be too small to reliably assess model mismatch, Bar-Shalom et al. [1, p.237] advocate combining test statistics from several time steps. Because the time series of $\tilde{y}_0 - H\mu_0$ values is an innovations sequence, random variables (6) from different time instants are mutually independent. Consequently, the sum of values from k different instants has a χ^2_{nk} distribution. The corresponding Bayesian p-test is to determine whether the sum of ϵ_0 values from k different instants lies in the tails of the χ^2_{nk} distribution. This is equivalent to the time-average NIS test of Bar-Shalom et al.

Now consider a test statistic based on the *posterior* predictive distribution. It follows from (4) that the posterior random variable

$$(\tilde{y}_1 - H\mu_1)' S_1^{-1} (\tilde{y}_1 - H\mu_1) \tag{13}$$

has a χ^2_n distribution. The corresponding Bayesian p-test is to determine whether the test statistic

$$\epsilon_1 = (y - H\mu_1)' S_1^{-1} (y - H\mu_1) \tag{14}$$

lies in the tails of the χ^2_n distribution. This is a new filter consistency test that, like the NIS test, can in principle be implemented as an online procedure because it does not require knowledge of the true state or of future measurements.

Comparing (12) and (14), it can be seen that the test statistics $\epsilon_r = \epsilon_0$ and ϵ_1 are both weighted sums of squares of the realised residuals $y - H\mu_1$, but with different weighting matrices. In Appendix A it is shown that the statistics are related by the inequality

$$\epsilon_1 \leq \epsilon_0. \tag{15}$$

Thus ϵ_1 is in the lower tail of χ^2_n whenever ϵ_0 is, and ϵ_0 is in the upper tail of χ^2_n whenever ϵ_1 is.

2.3 Augmented model approach

Consider now the following application of the augmented-model approach advocated by Kruschke [6]. In order to assess whether the nominal measurement

noise covariance is correct, the base model is augmented with a parameter τ that scales the measurement noise covariance; the measurement's sampling model is then $y \,|\, x, \tau \sim \mathrm{MVN}(Hx, \frac{1}{\tau}R)$. The base model is obtained as the special case $\tau = 1$.

In order to be able to obtain a posterior in terms of standard distributions, it is assumed that the prior distribution of τ is a gamma distribution and that $x \,|\, \tau \sim \mathrm{MVN}(\mu_0, \frac{1}{\tau}P_0)$. The prior for τ is denoted $\tau \sim \mathrm{gam}(a_0, b_0)$, a gamma distribution with location a_0 and scale b_0. Then, as derived in [7, p. 118], the marginal posterior of τ is

$$\tau \,|\, y \sim \mathrm{gam}\Big(a_0 + \frac{n}{2}, b_0 + \frac{2}{\epsilon_0}\Big). \tag{16}$$

If most of the posterior probability of τ is located far below the nominal value $\tau = 1$, then the nominal measurement noise covariance in the base model can be inferred to be too small to be consistent with the data. Similarly, a too-large noise covariance in the base model is indicated when $\tau = 1$ is far in the lower tail of the posterior gamma distribution.

There remains the question of choosing the parameters for the prior of τ. A conventional choice for a diffuse prior of a variance parameter is the scale-invariant improper density $p(\tau) \propto \frac{1}{\tau}$, which corresponds to the gamma with $a_0 \to 0$ and $b_0 \to 0$. The marginal posterior (16) then reduces to

$$\tau \,|\, y \sim \mathrm{gam}\Big(\frac{n}{2}, \frac{2}{\epsilon_0}\Big), \tag{17}$$

or equivalently, $\tau\epsilon_0 \,|\, y \sim \chi^2_n$. Appraising whether this posterior τ is far from the base model's nominal value $\tau = 1$ can be done by checking whether ϵ_0 lies in the tails of the χ^2_n distribution; this procedure is equivalent to the NIS test.

3 Tracking example

The performance of the online filter consistency tests is now illustrated with some simple examples with computer-generated data.

Consider first the example of [1, chap. 5], which has process model

$$x_{t+1} \,|\, x_t \sim \mathrm{MVN}\Big(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_t,\, q \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}\Big), \tag{18}$$

known initial state

$$x_0 \sim \mathrm{MVN}\Big(\begin{bmatrix} 0 \\ 10 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\Big), \tag{19}$$

and measurement sampling model

$$y_t \,|\, x_t \sim \mathrm{MVN}(\begin{bmatrix} 1 & 0 \end{bmatrix} x_t,\, r). \tag{20}$$

Computer simulations of a 100-step trajectory are made with states and measurements generated using noise intensity values of q and r that may differ from the filter model's nominal values $q_{\mathrm{F}} = r_{\mathrm{F}} = 1$. The test quantities ϵ_0 and ϵ_1 are computed at each time step (Figure 1), and the number of times they are outside the two-sided 95% χ^2_1 acceptance bounds $B = (0.001,\, 5.024)$ is counted (Table 1). In the results, neither test has a model mismatch detection rate better than 50%. Apparently, there is too little information in a single measurement to allow effective filter consistency assessment with these tests.

Table 1: Consistency tests for the example (18–20). Measurements are generated with the given q, r values and the filter having noise variance parameters $q = r = 1$ is run for 100 time steps.

q	r	$\%\{\epsilon_0 < B_1\}$	$\%\{\epsilon_0 > B_2\}$	$\%\{\epsilon_1 < B_1\}$	$\%\{\epsilon_1 > B_2\}$
1	1	2	3	7	0
10	1	0	27	1	2
1	10	2	27	4	2
0.1	1	2	0	7	0
1	0.1	3	0	11	0

Consider now time-averaged versions of the consistency tests applied to the same simulation data. Moving-window sums of $k = 5$ consecutive samples of ϵ_0 and ϵ_1 are compared to the two-sided 95% χ^2_5 acceptance bounds $B = (3.25\, 20.48)$. In the results (Table 2), the time-averaged NIS test, which uses ϵ_0, has fair (63% or better) detection rate in the cases where the process or measurement noise covariances are larger than in the nominal model. The new consistency test based on posterior predictive distributions, which uses

ϵ_1, has high (over 90%) detection rate in cases where the process or measurement noise covariances are smaller than in the nominal model. However, the new test also has fairly high (59%) false detection rate when the nominal model is correct.

Table 2: Time-averaged consistency tests for the example (18–20), based on a moving window of width $k = 5$.

q	r	$\%\{\sum \epsilon_0 < B_1\}$	$\%\{\sum \epsilon_0 > B_2\}$	$\%\{\sum \epsilon_1 < B_1\}$	$\%\{\sum \epsilon_1 > B_2\}$
1	1	2	0	**59**	0
10	1	0	**63**	13	0
1	10	0	**66**	9	2
0.1	1	2	0	**91**	0
1	0.1	14	0	**99**	0

Finally, consider a modification of the example so that there is more redundancy in the measurements. Let $n = 5$ independent measurements be made at each time step, so that (20) is replaced by

$$y_t \,|\, x_t \sim \mathrm{MVN}\Big(\begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} x_t,\, rI\Big). \tag{21}$$

The two-sided 95% χ^2_5 acceptance bounds are $B = (0.83,\, 12.83)$.

In the simulation results (Table 3, Figure 2) both tests have 90% or better model mismatch detection rate when the nominal measurement covariance is too small. When the nominal measurement covariance is too large, the new test has better detection rate (92%) than the NIS test (56%). Neither test has a detection rate better than 30% when only the process covariance is too large or too small.

4 Consistency in estimation of a scalar

As is done in [2], further insight into the consistency tests can be obtained by consideration of the elementary problem of estimating a scalar from n iid measurements. Consider measurments that are modelled as univariate normal with mean x and variance ρ, denoted $y_i \,|\, x \overset{\text{iid}}{\sim} \mathrm{N}(x, \rho)$. Let the prior

Table 3: Consistency tests for the example with $n = 5$ measurements per time step. Measurements are generated with the given q, r values and the filter having noise variance parameters $q = r = 1$ is run for 100 time steps.

q	r	$\%\{\epsilon_0 < B_1\}$	$\%\{\epsilon_0 > B_2\}$	$\%\{\epsilon_1 < B_1\}$	$\%\{\epsilon_1 > B_2\}$
1	1	2	3	6	2
10	1	0	28	1	2
1	10	0	**95**	0	**90**
0.1	1	2	0	4	0
1	0.1	**56**	0	**92**	0

be $x \sim \mathrm{N}(0, \pi_0)$. The posterior distribution of the state is then

$$x \,|\, y \sim \mathrm{N}(\tfrac{1}{\frac{\rho}{n\pi_0}+1}\bar{y}, \tfrac{1}{\frac{1}{\pi_0}+\frac{n}{\rho}}), \tag{22}$$

where $\bar{y} = \frac{1}{n}\sum y_i$. As shown in Appendix A, the prior predictive test quantity for this example is

$$\epsilon_0 = \frac{nv}{\rho} + \frac{1}{\frac{\rho}{n}+\pi_0}\bar{y}^2, \tag{23}$$

where $v = \frac{1}{n}\|y - \bar{y}\|^2 = \frac{1}{n}\|y\|^2 - \bar{y}^2$ is the empirical variance.

By examination of (23) it can be seen that ϵ_0 is large when the empirical variance is large (i.e. $v \gg \frac{\rho}{n}$) and/or the observations' mean is far from the (zero) prior mean (i.e. $|\bar{y}| \gg \sqrt{\frac{\rho}{n}+\pi_0}$). In other words, an excessively large ϵ_0 indicates that the model's ρ value is too small or the prior mean is wrong. An excessively small ϵ_0 means that ρ is too large and that the prior mean is correct.

As shown in Appendix A, the posterior predictive test quantity for this example is

$$\epsilon_1 = \frac{nv}{\rho} + \frac{1}{(\frac{\rho}{n}+\pi_0)(1+2\frac{\pi_0 n}{\rho})}\bar{y}^2. \tag{24}$$

Because the second term tends to zero as n grows large, it follows that when there is a large amount of data at the time step (i.e. $n \gg \frac{\rho}{\pi_0}$), the test quantity ϵ_1 is relatively insensitive to mis-modelling of the prior, and is too small (resp. too large) when ρ is too small (resp. large).

5 Conclusion

In this work, different procedures of Bayesian model criticism were explored as possible alternatives to the online NIS test of Kalman filter consistency. Three of the procedures turned out to be equivalent to the NIS test. We thus have three new interpretations of NIS: as a Bayesian p-test for the prior predictive distribution, as a nested-model parameter significance test, and as a test based on a weighted sum of squared residuals statistic.

A new NIS-like test was obtained from a posterior predictive Bayesian p-test. NIS and the new test statistic are weighted sums of squared residuals, but with different weighting matrices. In simulations, the new test outperforms the NIS test in the detection of undersized covariance parameters. Theoretical analysis of a basic estimation problem shows that the two tests are complementary, in that they detect different aspects of mis-modelling,

The present work extends only to linear Gaussian filtering. It is known that even when noise is additive zero-mean Gaussian, the effect of nonlinearity in the measurement function can be modelled as an additional term to the measurement covariance in the extended Kalman filter model [1, p. 385]. Thus, changes in the severity of nonlinearity may also lead to effects similar to those of filter covariance mis-modelling. A recent study on consistency tests for nonlinear filters is presented in [8].

References

[1] Yaakov Bar-Shalom, X.-Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2001.

[2] Richard G Gibbs. New Kalman filter and smoother consistency tests. *Automatica*, 49(10):3141–3144, 2013. doi: 10.1016/j.automatica.2013.07.013.

[3] Anthony O'Hagan and Jonathan Forster. *Kendall's Advanced Theory of Statistics: Volume 2B. Bayesian Inference*. Arnold, 2nd edition, 2004.

[4] George E. P. Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A*, 143(4):383–430, 1980.

[5] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd edition, 2004.

[6] John K Kruschke. Posterior predictive checks can and should be Bayesian. *British Journal of Mathematical and Statistical Psychology*, 66(1):45–56, 2013. doi: 10.1111/j.2044-8317.2012.02063.x.

[7] Karl-Rudolf Koch. *Introduction to Bayesian Statistics*. Springer, 2nd edition, 2007.

[8] Pavel Ivanov, Simo Ali-Löytty, and Robert Piché. Evaluating the consistency of estimation. In *International Conference on Localization and GNSS (ICL-GNSS)*, 2014.

A Derivations and proofs

A.1 Derivation of (2–4)

Introducing the measurement and predicted measurement noise variables w and $\tilde{w}$, the measurement models can be written as $y \,|\, x, w, \tilde{w} = Hx + w$ and $\tilde{y} \,|\, x, w, \tilde{w} = Hx + \tilde{w}$. The variables $x, w, \tilde{w}$ are mutually independent with $x \sim \mathrm{MVN}(\mu_0, P_0)$, $w \sim \mathrm{MVN}(0, R)$, and $\tilde{w} \sim \mathrm{MVN}(0, R)$. The full model is thus

$$\begin{bmatrix} x \\ \tilde{y} \\ y \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ H & I & 0 \\ H & 0 & I \end{bmatrix} \begin{bmatrix} x \\ \tilde{w} \\ w \end{bmatrix},$$

with

$$\begin{bmatrix} x \\ \tilde{w} \\ w \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mu_0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P_0 & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & R \end{bmatrix} \right).$$

Applying the formula for linear transformation of normal random variables and the notation of (1a) gives

$$\begin{bmatrix} x \\ \tilde{y} \\ y \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mu_0 \\ H\mu_0 \\ H\mu_0 \end{bmatrix}, \begin{bmatrix} P_0 & P_0H' & P_0H' \\ HP_0 & S_0 & HP_0H' \\ HP_0 & HP_0H' & S_0 \end{bmatrix} \right),$$

from which (3) is obtained by marginalisation of $\tilde{y}$. Then, applying the formula for conditioning of normal distributions and the notation (1) and (5) gives

$$\begin{bmatrix} x \\ \tilde{y} \end{bmatrix} \mid y \sim \mathrm{MVN}\Big(\begin{bmatrix} \mu_1 \\ H\mu_1 \end{bmatrix}, \begin{bmatrix} P_1 & P_1H' \\ HP_1 & S_1 \end{bmatrix}\Big),$$

from which (2) and (4) are obtained by marginalisation.

A.2 Proof that (7) = (8)

From (1d) it follows that $\mu_1 - \mu_0 = K(y - H\mu_0)$ and $y - H\mu_1 = (I - HK)(y - H\mu_0)$. Then

$$\begin{aligned}
&(\mu_1 - \mu_0)'P_0^{-1}(\mu_1 - \mu_0) + (y - H\mu_1)'R^{-1}(y - H\mu_1) \\
&\quad = \mathrm{tr}\big((\mu_1 - \mu_0)(\mu_1 - \mu_0)'P_0^{-1} \\
&\qquad + (y - H\mu_1)(y - H\mu_1)'R^{-1}\big) \\
&\quad = \mathrm{tr}\big(K(y - H\mu_0)(y - H\mu_0)'K'P_0^{-1} \\
&\qquad + (I - HK)(y - H\mu_0)(y - H\mu_0)'(I - HK)R^{-1}\big) \\
&\quad = \mathrm{tr}\big((y - H\mu_0)(y - H\mu_0)'(K'P_0^{-1}K \\
&\qquad + (I - HK)(I - HK)R^{-1})\big)
\end{aligned}$$

The result is then obtained by substituting (1d) and the identity $RS_0^{-1} = I - HK$.

A.3 Derivation of (10)

From (9) and the identity $RS_0^{-1} = I - HK$ we have

$$\begin{aligned}
\tilde{y} - H\tilde{\mu}_1 &= \tilde{y} - H(\mu_0 + K(\tilde{y} - H\mu_0)) \\
&= (I - HK)\tilde{y} - H(I - KH)\mu_0 \\
&= (I - HK)(\tilde{y} - H\mu_0) \\
&= RS_0^{-1}(\tilde{y} - H\mu_0).
\end{aligned}$$

Applying (3) then gives (10).

A.4 Proof of (15)

Using (8) and the fact that $S_1 - R = HP_1H'$ is non-negative definite, we have

$$\begin{aligned}\epsilon_0 &= (\mu_1 - \mu_0)'P_0^{-1}(\mu_1 - \mu_0) + (y - H\mu_1)'R^{-1}(y - H\mu_1) \\ &\geq (y - H\mu_1)'R^{-1}(y - H\mu_1) \\ &\geq (y - H\mu_1)'S_1^{-1}(y - H\mu_1) = \epsilon_1.\end{aligned}$$

A.5 Derivation of (22)

Let $m = \frac{\pi_0}{\rho}$, $H = \mathbf{1}$ (a column-vector of ones), and $R = \rho I$. Then

$$\begin{aligned}S_0 &= HP_0H' + R = \pi_0\mathbf{1}\mathbf{1}' + \rho I \\ K &= P_0H'S_0^{-1} = \pi_0\mathbf{1}'(\pi_0\mathbf{1}\mathbf{1}' + \rho I)^{-1} \\ &= \mathbf{1}'\Big(\mathbf{1}\mathbf{1}' + \frac{1}{m}I\Big)^{-1} \\ &= \mathbf{1}'\Big(mI - \frac{m^2}{1+mn}\mathbf{1}\mathbf{1}'\Big) \qquad \text{[Woodbury formula]} \\ &= \frac{m}{1+mn}\mathbf{1}' \\ \mu_1 &= \mu_0 + K(y - H\mu_0) = \frac{m}{1+mn}\mathbf{1}'y \\ &= \frac{m}{1+mn}n\bar{y} = \frac{1}{\frac{\rho}{n\pi_0} + 1}\bar{y} \\ I - KH &= 1 - \frac{m}{1+mn}\mathbf{1}'\mathbf{1} = \frac{1}{1+mn} \\ P_1 &= (I - KH)P_0 = \frac{\pi_0}{1+mn} = \frac{1}{\frac{1}{\pi_0} + \frac{n}{\rho}}\end{aligned}$$

A.6 Derivation of (23)

$$\begin{aligned}
\epsilon_0 &= y' S_0^{-1} y = y'(\pi_0 \mathbf{1}\mathbf{1}' + \rho I)^{-1} y \\
&= \frac{1}{\pi_0} y'(\mathbf{1}\mathbf{1}' + \frac{1}{m} I)^{-1} y \\
&= \frac{1}{\pi_0} y' \Big(mI - \frac{m^2}{1+mn} \mathbf{1}\mathbf{1}' \Big) y \\
&= \frac{1}{\pi_0} \Big(m\|y\|^2 - \frac{m^2 n^2}{1+mn} \bar{y}^2 \Big) \\
&= \frac{1}{\pi_0} \frac{mn(1+mn)(v+\bar{y}^2) - m^2 n^2 \bar{y}^2}{1+mn} \\
&= \frac{1}{\pi_0} \Big(mnv + \frac{mn}{1+mn} \bar{y}^2 \Big) = \frac{nv}{\rho} + \frac{1}{\frac{\rho}{n} + \pi_0} \bar{y}^2
\end{aligned}$$

A.7 Derivation of (24)

$$\begin{aligned}
S_1^{-1} &= \frac{1}{\pi_0} \Big(\frac{1}{1+mn} \mathbf{1}\mathbf{1}' + \frac{1}{m} I \Big)^{-1} \\
&= \frac{1}{\pi_0} \Big(mI - \frac{m^2}{1+2mn} \mathbf{1}\mathbf{1}' \Big) \\
y' S_1^{-1} y &= \frac{1}{\pi_0} \Big(m\|y\|^2 - \frac{m^2 n^2}{1+2mn} \bar{y}^2 \Big)
\end{aligned}$$

$$\begin{aligned}
y' S_1^{-1} H \mu_1 &= \frac{m}{1+mn} y' S_1^{-1} \mathbf{1}\mathbf{1}' y \\
&= \frac{1}{\pi_0} \frac{m}{1+mn} \Big(m - \frac{m^2 n}{1+2mn} \Big) y' \mathbf{1}\mathbf{1}' y \\
&= \frac{1}{\pi_0} \frac{m^2 n^2}{1+2mn} \bar{y}^2
\end{aligned}$$

$$\begin{aligned}
\mu_1' H' S_1^{-1} H \mu_1 &= \frac{m^2}{(1+mn)^2} y' \mathbf{1}\mathbf{1}' S_1^{-1} \mathbf{1}\mathbf{1}' y \\
&= \frac{1}{\pi_0} \frac{m^2}{(1+mn)^2} \Big(mn - \frac{m^2 n^2}{1+2mn} \Big) y' \mathbf{1}\mathbf{1}' y \\
&= \frac{1}{\pi_0} \frac{m^3 n^3}{(1+mn)(1+2mn)} \bar{y}^2
\end{aligned}$$

$$
\begin{aligned}
\epsilon_1 &= (y - H\mu_1)' S_1^{-1} (y - H\mu_1) \\
&= y S_1^{-1} y - 2y' S_1^{-1} H\mu_1 + \mu_1' H' S_1^{-1} H\mu_1 \\
&= \frac{1}{\pi_0}\Big(m\|y\|^2 - \frac{m^2 n^2 (3+2mn)}{(1+mn)(1+2mn)}\bar{y}^2\Big) \\
&= \frac{1}{\pi_0}\Big(mnv + \frac{mn}{(1+mn)(1+2mn)}\bar{y}^2\Big) \\
&= \frac{nv}{\rho} + \frac{1}{(\frac{\rho}{n} + \pi_0)(1 + 2\frac{\pi_0 n}{\rho})}\bar{y}^2
\end{aligned}
$$

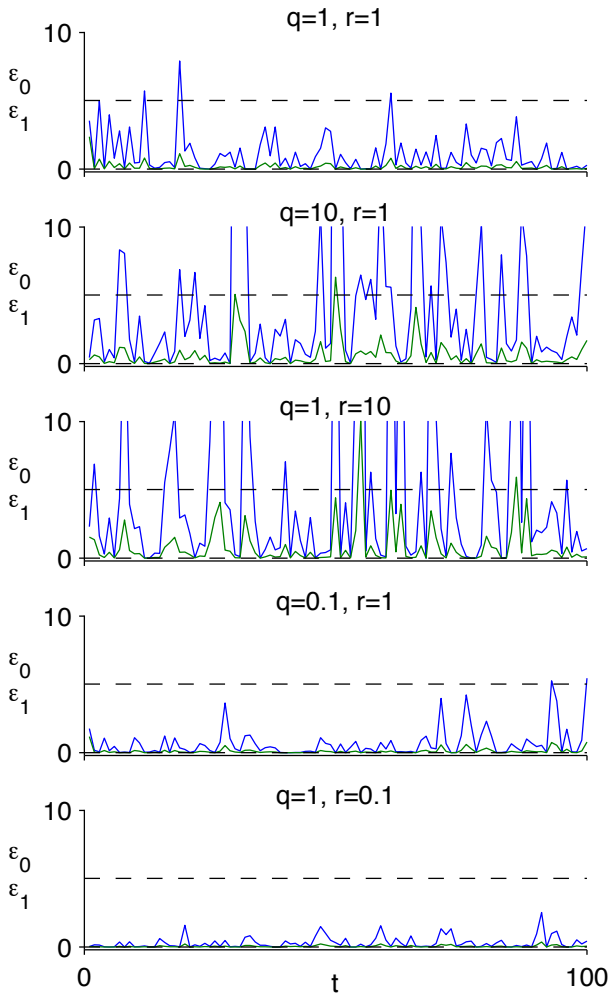

Figure 1: Time series for the simulation in Table 1. Solid lines are the test statistics ϵ_0 and ϵ_1, with $\epsilon_0 \geq \epsilon_1$; dashed lines are the two-sided 95% χ^2_1 acceptance bounds.

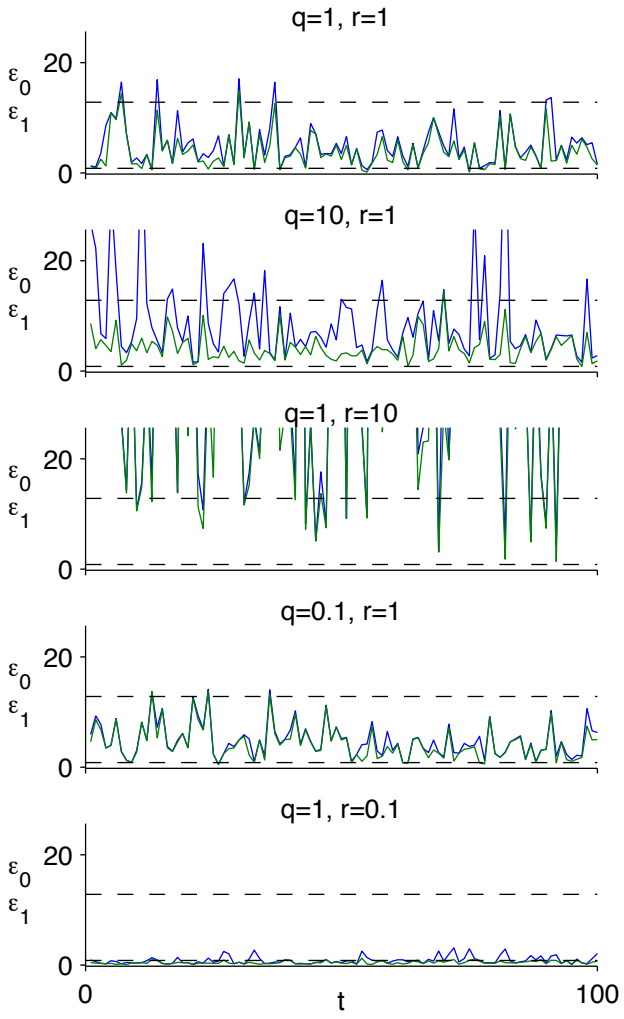

Figure 2: Time series for the simulation in Table 3. Solid lines are the test statistics $\epsilon_0 \geq \epsilon_1$; dashed lines are the two-sided 95% χ^2_5 acceptance bounds.